PB-229 566

THE STORAGE AND RETRIEVAL OF DATA FOR WATER QUALITY CONTROL

Richard S. Green

Public Health Service
Washington, D. C.

1964

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

PB 229 566

The Storage and Retrieval of Data for Water Quality Control

By RICHARD S. GREEN, Chief, Basic Data Branch

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
Public Health Service
Division of Water Supply
and Pollution Control
Washington, D.C. 20201

The material in this publication was used as the basis of a presentation at the Symposium on Environmental Measurements, September 5, 1963, Sheraton-Gibson Hotel, Cincinnati, Ohio.

8/18

U.S. GOVERNMENT PRINTING OFFICE, Washington, D.C. – 1964

Foreword

This system for the storage and retrieval of data for water quality control was developed from ideas brought together in a brief informal conference held in the Basic Data Branch, Division of Water Supply and Pollution Control, in August 1961.

Although not present at that meeting, the thoughts and suggestions of a few State officials who had been concerned about this problem were contributed by Public Health Service personnel familiar with their views. Operating procedures in Indiana, New York State, and Pennsylvania were especially helpful. The original concepts brought out at this meeting have been adjusted almost continuously over the ensuing months as their full potential became apparent.

Assistant Sanitary Engineer (R) Clarence Tutwiler, who worked on the project from March of 1962, has been responsible for the coding and card formats needed to implement this system, and for the electronic computer programing required. Mr. Tutwiler's great skill in adjusting, and readjusting, the storage and retrieval procedures has been equaled only by his patience as he dealt with the expanding demands of his colleagues.

Contents

	Page
Introduction	1
The Current Water Quality Control Data Handling Situation	3
Scope of the System	3
The System Design	4
Storage Procedure	5
Location Cards I, II	5
Location Card III	11
Location Card IV	11
Location Card V	13
Parameter Card	13
Deleting and Changing Data	15
Use of Data From Other Systems	17
Parameter Abbreviation Card	18
Retrieval Routine	18
Parameter Control Card	21
Station Control	
Card I	21
Cards II, III, IV	23
Date and Type-of-Station Control	24
Selection of Output Format	26
References	27
Appendix I—Storage of Data	27
Appendix II—Retrieval of Data	29
Appendix III—Card Format and Coding	30–36

Introduction

Much public money is spent measuring the effect of environmental changes on man and of man's activities on his surroundings. It is incumbent on those of us in responsible positions to make sure that maximum value accrues from these expenditures.

As pertinent data become more widely needed by the groups and agencies involved in air and water pollution control, greater importance must be placed on uniformity of sampling and analytical procedures, and on the accessibility of reliable data from all sources. Data now published by local, State, or Federal agencies cannot easily be presented in uniform format because quality parameters, levels of concentration, objectives, and requirements differ. More often than not, an agency secures data for one particular study, giving little thought to possible use of the information by others, for different purposes. Thus many data reports are of limited use, requiring costly and time-consuming extraction, relisting, and analysis to be applied to situations other than those of first concern. Too often data are considered to have served their purpose once they become a part of a specific study on a particular situation—and appear in the appendix of a report. Moreover, a large body of valuable data never appears in print at all, but remains in inaccessible files until discarded.

Wherever acquired data are likely to have lasting value, serious thought should be given to some system of storage and retrieval wherein potential users of the data—

1. Can obtain the information they need in the form they need it;
2. Will be assured that all reliable information, wherever produced, is included in the material they request, and that all areas of interest are covered;
3. Will not be bothered with data they do not need; and
4. Can get this service at minimum cost and with reasonable speed.

Any such system, to be workable, must possess three important characteristics. It must be—

1. *Comprehensive*—have the ability to handle all possible quality and related parameters, both those now in use and those which may be significant in the future;

2. *Flexible*—be able to take into account geographic and environmental differences; and

3. *Simple*—be relatively easy to use, and be within reasonable cost range.

Designing a system of this scope involves many problems. Most have been encountered, and solved, in a procedure for storing and retrieving data for water quality control, devised for general use in the operations of the Division of Water Supply and Pollution Control of the Public Health Service. A description of the elements of that system, which has been named STORET, will bring out many of the basic principles involved. The system is both comprehensive and flexible and, it is believed, as simple as the requirements will permit.

The Current Water Quality Control Data Handling Situation

There can be little doubt that the present confused status of data handling across the country in the field of water pollution control has contributed significantly to the tardiness in coming to grips with many pollution situations.

Walter Lyon, in his presentation before the Symposium on Water Quality Measurement and Instrumentation at Cincinnati in August of 1960, pointed out that in a 75-mile stretch of one eastern seaboard river the following groups had collected or were collecting water quality data:

3	Federal agencies
1	interstate agency
5	State agencies, representing 3 States
20	industries
7	sewage works
7	water works
3	consulting firms
46	

Doubtless similar situations could be found in most of our active river systems. Could some of this work have been avoided if there had been an easy way to pool the data on this stream? Mr. Lyon certainly was of this view, for he went on in his talk to stress the need for more uniformity in producing and processing data, and he pointed out some of the difficulties to be overcome.

While a completely uniform method of handling data in the water field may not be practicable in the near future, the Division of Water Supply and Pollution Control does have a compelling responsibility to see that its own data activities are logically brought together, and a real interest in finding the most efficient way of using data produced by others. This Division is now producing data far in excess of anything predicted even a few years ago. In all parts of the country many different comprehensive projects and related activities are developing fundamental information which, for maximum use, must be readily available to all interested persons. Significant internal problems in applying this data-handling system to existing operations must be faced, but they are not insurmountable, and the dividends which will accrue when uniformity prevails will more than outweigh them. By this unified approach to data storage and retrieval today, great difficulties will be avoided in the future. Furthermore, the experience gained through the use of this particular system should enable the Public Health Service to contribute more effectively to the more general development and application of uniform data-handling methods which so many in the water field believe is necessary.

Scope of System

As will be apparent later in this presentation, the size and complexity of this data-handling problem dictates the use of electronic computers with their great storage capacity and ready access to selected items. Two major concepts are applied uniformly throughout the country regardless of the computer equipment or programing technique utilized: (1) A single procedure for the identification of point locations pertinent to the data, be they water quality sampling points, points of waste discharge or of water intake, or any other

locations for which data are to be secured; (2) a single coding system for the identification of specific parameters of water quality or other particular items of interest in our field, e.g., data on flow, precipitation, and the like.

Location Code.—The location code is suitable for national use, and permits the retrieval of data in the hydrologic order that is so desirable for studies of basin problems. Since the system is complete and "open ended," it is possible to identify any point on any stream by it. Once a data point is properly identified by the coding outlined, this "label" remains with that data point indefinitely. The location code is not, at present, completely adapted to points in estuarial waters where interlocking channels cannot easily be fitted into the concept, nor to large open water bodies. It was deemed unwise, however, to delay the application of other features of the plan on this account. Some form of coordinate location system will probably be used for such points.

Parameter Code.—One of the most troublesome features in the handling of data, particularly water quality data, for wide geographic areas is that we find, first, a large number of different kinds of data of concern to us. Also, in any given type of measurement there are usually widely varying limits to the values reported from place to place: Chlorides in New England streams are usually very low, but very high in some streams of the Southwest. If we provide a fixed field size on a punchcard to accommodate the maximum values to be reported, we waste columns when less than that number is used. Furthermore, if we make room on a fixed field card for several parameters of data, not all of which are reported in each use of that card type, we waste still more columns. The availability of magnetic tape data storage enables us to overcome these difficulties. A parameter code has been adopted which will handle up to 100,000 parameters. Blocks of numbers have been assigned for specific parameter groups, leaving wide areas unfilled for future determinations and related data of interest (*1*).

In this system, the code for any individual parameter is the same wherever the data are secured, and any potential user of the stored information can call for specific kinds of data through use of the proper code number. A special comment is necessary with respect to the handling of biological data, the plankton, benthos, and fish. The number of individual entities here, i.e., species of organisms or fish, which may need to be reported and retrieved from storage for a given station, the need to record supplementary data about each entity, and the need to store and retrieve these data in associated groups, call for a different method of storage and retrieval than proposed for all other types of data. This is discussed in greater detail on page 18.

The problem of variable number of digits in the value for any given parameter is handled by limiting the reported value to four significant figures, with the decimal point coded as the applicable exponent of 10.

Statistical Analyses of Data.—Early in the design of this system it was decided that no attempt would be made to build in any procedures for statistical analysis of the stored data since this would only complicate the job of storage and retrieval. However, since the output from the system can be in the form of data on tapes in prearranged order as requested by the user, subsequent statistical processing is a simple matter.

Status of the System.—The programing of all routines involved in this data storage and retrieval system has been completed. Index coding and stream mileage measurements are being undertaken at the present time by several of the comprehensive river basin projects within the Division of Water Supply and Pollution Control in preparation for full-scale use of the system.

The System Design

The approach described in this presentation is built around the Minneapolis-Honeywell 400 Computer, though the system can be adapted to any equivalent computer of other manufacture. Since data will be obtained from many sources and compiled originally for different reasons, the storage system for accepting these data has been

made flexible. In general, the data to be handled are of two major types: first, water quality and similar data such as flow data which can be expected to vary, and therefore will be obtained in a repetitive manner; second, data which are not likely to change. Examples of this latter type are information about location of the point in question and about such items as treatment processes for waterworks installations. To serve the user, the former type of data must be obtainable from the storage system in a variety of ways, according to point of sampling, the date, specific parameters available, and similar variations. The latter, on the other hand, can be stored as a single block of information for a single location and extracted as a unit.

The details of this system are presented in terms of the various punched cards, shown in appendix III, used in storing and retrieving the data. The storage and retrieval mechanism and examples of the various ways in which the system can be used are brought out as these cards and their use are described. Field engineers, hydrologists, and others who use this data-handling system need not, of course, have an intimate knowledge of all these procedures. Nevertheless, it has been our experience that data-processing "customers" can more quickly reach an understanding with computer specialists, and the results will generally be more satisfactory, if the customer knows some of the basic tools of the specialist.

Storage Procedure

Before information such as water quality data can be stored on magnetic tape or retrieved in any logical order,* the point of sampling must be suitably described and coded into cards. The same holds for water intakes, sewage outfalls, gaging stations, and the like. Location cards I, II, III, IV, and V are used to locate and describe a station on the stream. (Note: throughout the following sections, the word station will refer to a point on a stream whether it be a sampling point, sewage outfall, or water intake. See item 4 under location card IV, page 11, for types of stations.)

*See app. I and II, "Storage of Data" and "Retrieval of Data" for the order used in the storage and retrieval of the information.

LOCATION CARDS I AND II:

Figures 1 and 2, appendix III, represent location cards I and II. These cards are used to store the location code for a station on tape so that the data for that station can be retrieved in hydrologic sequence.

Columns 1–6 of location card I are coded with a six-digit station code assigned serially by State. Thus, for example, each station in Alabama will be assigned an individual code number, the first two digits being 01. This code number is used to specify the station when storing or retrieving water quality and related data. Up to 10,000 stations per State can be coded in this fashion and this should suffice for most States; however, in the event that more than 10,000 stations are located in a State, an additional block or blocks of code numbers may be assigned to the stations in that State. Thus, for example, if Alabama had 15,000 stations, each station could have an individual code number beginning with either 01, the normal 2-digit code for Alabama, or 58.

Columns 7–76 of location card I and 14–45 of location card II are used to store the station location code which places the station on the river systems of the United States. The use of the location code is outlined below, and more completely discussed in a separate document. (*2*)

Location Coding by River Systems.—The water from almost every stream in the United States eventually flows either into an ocean at a point somewhere on the U.S. coastline, across the border at some point into Canada or Mexico, or into a landlocked lake or one of the Great Lakes at some point on the coastline of the lake. Each of these possible points is the mouth of a river system and each of the river systems in the United States is given a seven-digit code number. The first four digits of this code number locate the position of the mouth of the river system within a major basin and minor basin. Plate 1 shows the 15 major river basins of the United States. The last 3 digits of the code number differentiate between

PLATE 1

the river systems whose mouths are located in the same major basin and minor basin, thus providing for up to 999 primary river systems in each of these "terminal" minor basins. The actual discharge streams in each of these river systems are called "terminal streams"; also, the major basin and minor basin *in which the mouth of the terminal stream is located* will be referred to as the terminal major basin and terminal minor basin. This should not be confused with the major basin and minor basin in which the station itself is located, which may be different. For example, although station 2 shown in plates 2 and 3 is located in major basin 50 and minor basin 14, it will have 5110 as the first four digits of its location code since 5110 is the code number of the *terminal* major and minor basins of the river system in which the station is located. As another example, a point on the Ohio River at Cincinnati, although itself located in the Ohio River major basin, has as its terminal major basin the southwest-lower Mississippi. Terminal minor basin boundaries are being carefully defined on topographic sheets to permit the assignment of primary river system identification numbers.

The seven-digit code which designates a river system is assigned as the first seven digits of the location code of all stations located on any of the streams within that river system.

The next six digits of the code, columns 14–19 of location card I, represent the miles from the mouth of the river system to the station, if the station is located on the main stream of the river system, or to the confluence of the main stream with a tributary above which confluence point the

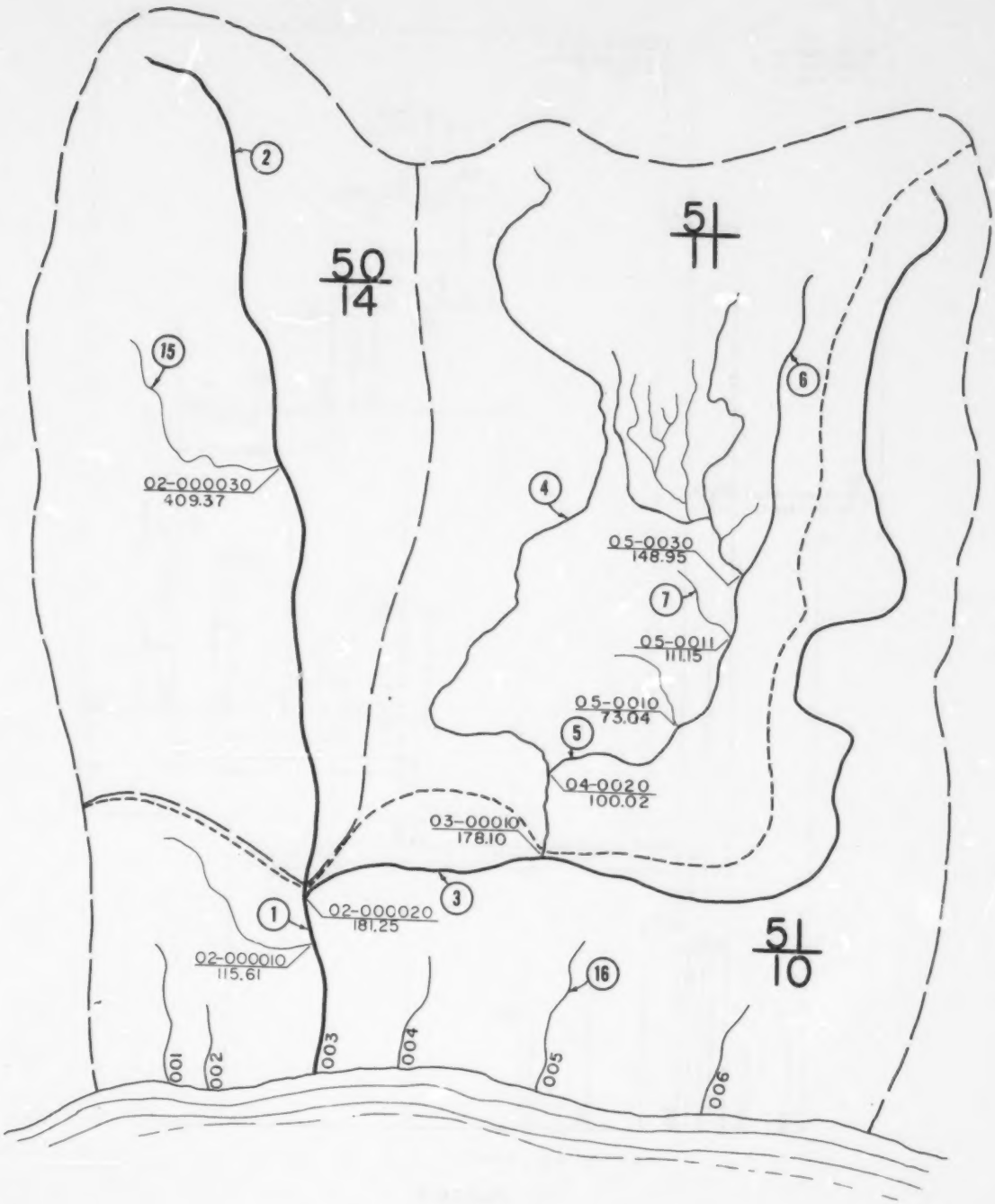


PLATE 2

River Systems in Hypothetical Terminal Basin 51–10

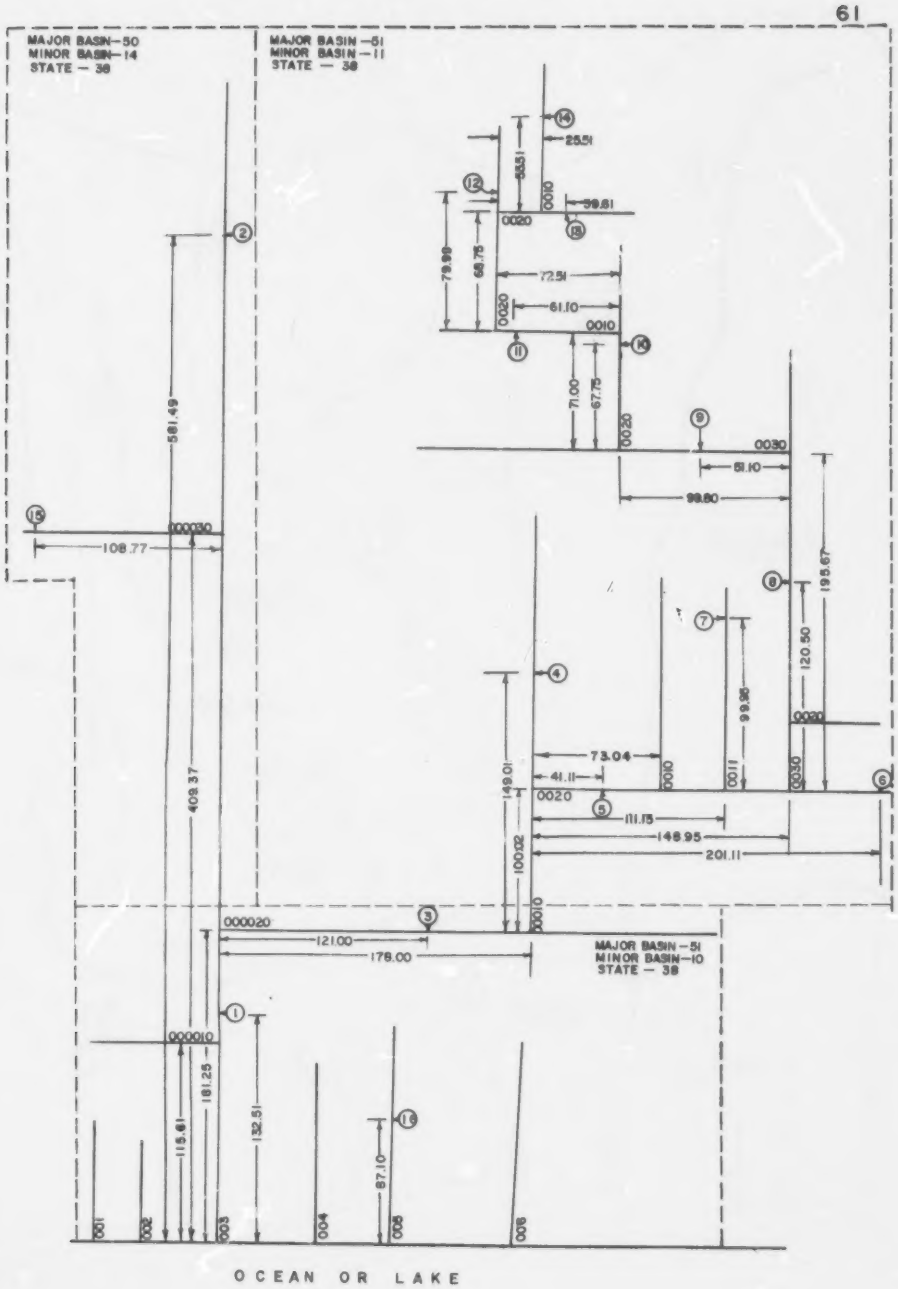


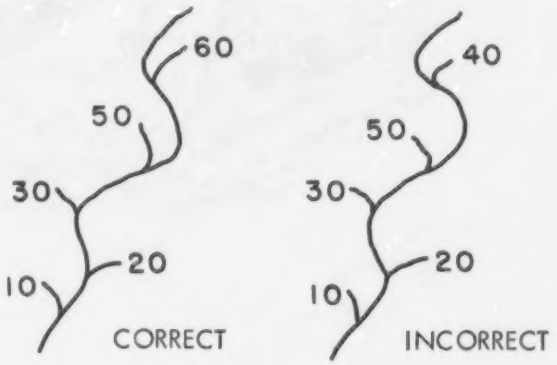
PLATE 3

Coding of River Systems in Hypothetical Terminal Basin 51–10

station is located. In this case and all of the following cases, after the last mileage to a station has been coded, all columns that are not used up to column 77 of location card 1 are left blank and if location card II is not needed to complete the code, it will be omitted. All mileage figures are shown and punched to hundredths of a mile. It is not possible, of course, nor is it necessary, for locations along streams to be measured this accurately on maps of reasonable scale. This degree of fineness in reporting, however, does make possible proper sequential placement on the storage tape of closely located or adjacent points.

Level Coding—All tributaries to the main stream of a river system are assigned a level II index code as follows: The main stem, using suitable maps,* is followed upstream and as each tributary is reached it is assigned a code number. The code numbers for tributaries entering from the left will be odd and for those entering from the right, the code numbers will be even. The code numbers will be assigned consecutively: i.e., if a tributary that enters the main stream is assigned the value N, the next tributary upstream is assigned the value N+10 or N+20, depending on whether it enters from the opposite or same side as the previous tributary. Thus, the stream system would be coded as shown in the left diagram, not the right diagram.

The number thus assigned to the tributary is then coded, right justified, into columns 20–24 of of the location card I, unused left-hand columns in 20–24 being filled with zeros. Column 25 will be coded with a zero unless the tributary on or above which the station is located was overlooked and omitted in an earlier coding of the river system. In this case the level II index code of the station is given by the level II index code of the previous downstream tributary with column 25 being assigned a value between 1 and 9, such that the value is even if the stream enters from the right and odd if it enters from the left.



The next field on location card I, Miles on Second Level, represents the miles from the mouth of the level II stream to the station, if the station is located on the level II stream, or to the confluence of the level II stream with the tributary above which confluence the station is located.

If the station has not yet been reached, i.e., if it is not on the level II stream, then the proper level III code is selected. This code is assigned to the tributaries of the level II main stream by the same procedure as the level II code was assigned to the tributaries of the terminal stream of the river system, and again the mileage to the station or to next branch is coded. This procedure is repeated for the fourth level, fifth level, etc., until the station in question is completely located. All streams above the seventh level are coded by using location card II.**

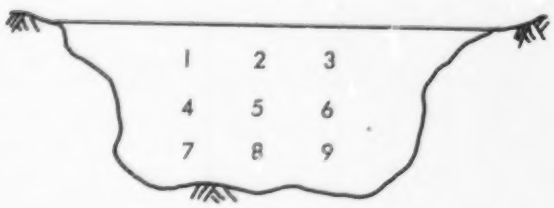
* In applying this code, the scale and type of map to be used are important. Wherever available, USGS topographic sheets of scale 1:24,000 or 1:62,500, with the brown contour lines and green vegetation overlays omitted, and with roads and other culture details suppressed in intensity, are usually best for this purpose. Since the degree of coverage at any given scale varies in different parts of the United States, coding in some areas will have to be at a scale of 1:250,000. Plate 4 shows the type of map that has been found most useful in this work.

**It is frequently necessary to locate points on the shore lines of terminal basins or large lakes such as the Great Lakes. This can be done by assigning, for shore line points, a two-digit minor basin number not already assigned to a minor basin in the major basin under consideration. The shore line itself can be considered as if it were a base river and will be given a three-digit number, 001, to be placed in columns 11–13 of location card I. Using an arbitrary point as zero mileage, all other points along the shore can be placed with respect to that zero point and properly logged in columns 14 through 19 of location card I.

SUSQUEHANNA
LOWER SWATARA
RIVER
Highspire
High Spire PO
Marsh Run
Hook Island
Fritz Island
Penbrook
Progress
Harrisburg
Penna Turnpike
Toll Road
U S MIL RES
Frogtown
Fishing Creek
Salem Ch
NEW
Plainfield
Yocumtown
Goldsboro
(Etters PO)
Wickersham Hill
NEWBERRY
Lewisberry
Redland Friends Ch
Cem
Neiman Hill
St Paul Ch
Township Sch
Newberrytown
Woodside
Cann-ah-on YWCA Camp
Susquehanna
Fishing
Bennett
Run
Creek
Dauphin
York Co

PLATE 4

The number of the level on which the station is located is then coded into columns 77 and 78 of location card I. Column 79 of this card may be coded with a number that locates the relative point of sampling within the cross sectional area of the stream, in accordance with the diagram below:



Cross sectional numbering system, looking upstream.

Thus, for example, if the point of sampling is on or near the left bank of a stream, looking upstream, the number 1 will be coded into column 79. A more precise method of locating and defining cross-section points is available through the use of horizontal and vertical coordinates, described in the section on the parameter abbreviation card, page 18. More detailed descriptive information about the sampling point is stored into location card V. Column 80 in all of the cards in appendix III is used by the computer to identify the specific type of card; the proper punch code for each of the cards is shown in the format of the card.

In order to illustrate the coding procedure for the location cards, a hypothetical terminal minor basin with several river systems is shown schematically in plate 3. Plate 2, which depicts the streams in these same hypothetical basins, has been partially coded to show how maps are marked. The index, preceded by a two-digit designation of the stream level, appears above the horizontal line. The mileage to the stream confluence is shown below the horizontal line. The codes of the location cards for 16 stations located in this area are given in table 1. Station 7 is a station located on a stream which was omitted in the original coding, because an earlier edition, faulty map was used.

LOCATION CARD III:

Location card III, figure 3, appendix III, has five fields to be coded; they are—

1. Station code serial—Same six-digit code as in location card I, columns 1–6.

2. Agency—The code number of the agency or organization collecting the data for this station.

3. State name—State in which station is located.

4. Major basin name—Name of major basin in which the station is located. This is not necessarily the terminal major basin. It is the actual major basin in which the station is located.

5. Minor basin name—Name of minor basin in which the station is located. This is not necessarily the terminal minor basin. It is the actual minor basin in which the station is located.

Since the purpose of fields four and five, above, and of fields two and three on location card IV, is to show the actual location of the station in every data printout, these fields can be used in any way that will best accomplish that goal. Other words, for example, can be substituted for these basin names without in any way interfering with the storage or retrieval routines.

LOCATION CARD IV:

Location card IV, figure 4, appendix III, has four fields to be coded; they are:

1. Station code serial—As before.

2. Minor basin name—(Continued from last card).

3. Location name—Name of station.

4. Type of station—Used to differentiate between the various types of stations, and coded—

002—if station is a point where water quality data is collected (i.e., a water quality sampling station).

004—if station is a point where stream flow is recorded (i.e., a gaging station).

006—if station is both a water quality sampling and gaging station.

010—if station is a municipal water intake where data on raw water is collected.

TABLE 1

Code of Location Cards I and II for Stations 1–16

STATION	CARD COLUMN								
	0	10	20	30	40	50	60	70	80
	1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9	0
1	3 8 0 0 0 1 5 1 1	0 0 0 3 0 1 3 2 5 1						0 1 2	1
2	3 8 0 0 0 2 5 1 1	0 0 0 3 0 5 8 1 4 9						0 1 3	1
3	3 8 0 0 0 3 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 2 1	0 0				0 2 4	1
4	3 8 0 0 0 4 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 4 9	0 1			0 3 1	1
5	3 8 0 0 0 5 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 0 4 1 1	1		0 4 2	1
6	3 8 0 0 0 6 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 2 0 1 1	1		0 4 7	1
7	3 8 0 0 0 7 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 1 1 1	5 0 0 1 1 0 9 9 9 5		0 5 9	1
8	3 8 0 0 0 8 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 2 0 5 0		0 5 8	1
9	3 8 0 0 0 9 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 5 1 1 0	0 5 5	1
10	3 8 0 0 1 0 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 9 9 8 0 0	0 2 0 7 7 5 0 7 6	1
11	3 8 0 0 1 1 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 9 9 8 0 0	0 2 0 7 1 0 0 0 8 9	1
*	3 8 0 0 1 1 5 1 1	0 0 0 3 0 0 1 0 6 1	1 0						2
12	3 8 0 0 1 2 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 9 9 8 0 0	0 2 0 7 1 0 0 0 9 5	1
*	3 8 0 0 1 2 5 1 1	0 0 0 3 0 0 1 0 7 2	5 1 0 0 2 0 7 9 9 9						2
13	3 8 0 0 1 3 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 9 9 8 0 0	0 2 0 7 1 0 0 1 0 7	1
*	3 8 0 0 1 3 5 1 1	0 0 0 3 0 0 1 0 7 2	5 1 0 0 2 0 6 8 7 5	0 0 2 0 3 9 6 1					2
14	3 8 0 0 1 4 5 1 1	0 0 0 3 0 1 8 1 2 5	0 0 0 0 2 0 0 1 7 8	0 0 0 0 0 1 0 1 0 0	0 2 0 0 2 0 1 4 8 9	5 0 0 3 0 1 9 5 6 7	0 0 3 0 0 9 9 8 0 0	0 2 0 7 1 0 0 1 1 9	1
*	3 8 0 0 1 4 5 1 1	0 0 0 3 0 0 1 0 7 2	5 1 0 0 2 0 6 8 7 5	0 0 2 0 2 5 5 1 0 0	1 0 5 5 5 1				2
15	3 8 0 0 1 5 5 1 1	0 0 0 3 0 4 0 9 3 7	0 0 0 0 3 0 0 1 0 8	7 7				0 2 1	1
16	3 8 0 0 1 6 5 1 1	0 0 0 5 0 0 8 7 1 0						0 1 1	1

*Second card for station (used to code all levels above seventh level).

020—if station is a water intake for a municipal water treatment plant where data on finished water is collected. (A description of water treatment can be stored for this type of station.)

040—if station is a municipal sewage outfall where quality data on the sewage is collected. (A description of sewage treatment can be stored for this type of station.)

100—if station is a water intake other than a municipal intake where data on the raw water is collected. (A description of water treatment can be stored for this type of station.)

200—if station is a waste outfall other than a municipal outfall where quality data on the effluent is collected. (A description of waste treatment can be stored for this type of station.)

400—open, for future use.

Additional types of stations can be included in the system, if necessary.

If data for both raw water quality and finished water quality for a municipal treatment plant are to be stored, two stations would be defined at the intake point. These two stations would then be assigned individual station code serial numbers, and one of the descriptive paragraph cards (see location card V) can also be punched with the treatment information. One station would be assigned station code 010 while the other would be assigned station code 020. The other information contained in the location cards for both stations would be identical.

LOCATION CARD V:

Location card V, figure 5, appendix III, is used to store a paragraph describing the station. This paragraph can contain up to 792 characters, since up to 11 cards of this type can be coded per station, with one line of the paragraph per card. This paragraph, for example, can be used to give longitude and latitude of the station and describe the drainage basin, if the station is a sampling point. If the station is a type of station designated by 020, 040, 100, or 200 (see location card IV), the treatment can be described on one of these cards. Other significant information can be included also.

Plate 5 shows a form used for coding location data.

PARAMETER CARD:

The card used for the storing of water quality and other variable data is the parameter card, the format of which is given in figure 6, appendix III.

In this card the station code field is coded with the 6 digit serial code of the station (see location card I). The date field for this card is coded with the date of the sample, i.e., year, month, and day. The field for time of year, for a composite sample, or hour and minute of a grab sample, can be coded in three ways:

1. If the sample is a composite sample, the year, month, and day of the final sample is coded into the field.

2. If the sample is not a composite sample but the time of day when the sample was taken is given, the hour (based on 24-hour day) and minute (based on 60-minute hour) is coded into columns 13–16. Columns 17–18 are then left blank.

3. If the sample is not a composite sample and the time of day of the sample is not given, columns 13–18 are left blank.

The next five fields (quality parameters 1–5) in the parameter card are individually coded with:

1. The parameter code, a five-digit code assigned individually to the different parameters (*1*).

2. Four significant digits of the value of that parameter, assuming the decimal point to be at the left of all significant figures.

3. An exponential value to give, in the printout, the proper location of the decimal point for the four significant figures of that parameter. The first exponential digit is 0 if exponential value is negative and 1 if exponential value is positive; the second digit is the exponential value itself.

4. Remark code—Up to 14 different remarks can be coded in this column. If no remark is needed, the column is left blank. Typical remarks that might be needed, and their corresponding codes are as follows:

STATION LOCATION DATA

CARD 1

STATION DESIGNATION — SERIAL NUMBER (1-6)

TERMINAL MAJOR BASIN NAME (7-8)

TERMINAL MINOR BASIN NAME (9-10)

RIVER SYSTEM NUMBER WITHIN TERMINAL MINOR BASIN (11-13)

LEVEL	INDEX	MILES
1		(14-19)
2	(20-25)	(26-31)
3	(32-36)	(37-41)
4	(42-45)	(46-50)
5	(51-54)	(55-59)
6	(60-63)	(64-68)
7	(69-72)	(73-76)

LAST LEVEL USED (77-78)

LOCATION IN CROSS SECTION (79) [1] (80)

CARD 2 REPEAT COLUMNS 1-13 ABOVE

LEVEL	INDEX	MILES
8	(14-17)	(18-21)
9	(22-25)	(26-29)
10	(30-33)	(34-37)
11	(38-41)	(42-45)

COLUMNS 46-79 BLANK [2] (80)

CARD 3

SERIAL NUMBER (1-6)

AGENCY NAME (7-12)

STATE NAME (13-28) (16 LETTERS AND SPACES)

MAJOR BASIN NAME WERE STATION IS LOCATED (29-52) (23 LETTERS AND SPACES)

MINOR BASIN NAME WHERE STATION IS LOCATED (53-79) (27 LETTERS AND SPACES) [3] (80)

CARD 4

SERIAL NUMBER (1-6)

MINOR BASIN NAME WHERE STATION IS LOCATED (CONT.) (7-19) (13 LETTERS AND SPACES, IF NEEDED)

LOCATION NAME (20-51) (32 LETTERS AND SPACES)

TYPE(S) OF STATION

(61-63)

BLANK (64-79) [4] (80)

REMARKS:

PHS 8-63 FOR ILLUSTRATION ONLY

PLATE 5

Punch	Printed output	Remarks
Blank	Blank	
0	J	value not accurate.
1	K	actual value is known to be less than value given.
2	L	actual value is known to be greater than value given.
3	M	
4	N	
5	O	
6	P	
7	Q	
8	R	
9	#	
2,8 *	$	
.	*	
@	''	

*Double punch 2 and 8 in same card column.

Additional remarks will have codes assigned to them as dictated by experience in the use of the system.

If the significant digit columns of a quality parameter field are left blank, this field is skipped and is not stored by the storage program. This option is useful when the same analyses are to be repeated on many samples. In this case the parameter cards can be gang punched with selected parameter codes previous to the value coding, thus making the parameter card a fixed field card and therefore easier to code.

A second option incorporated in the storage and retrieval program is useful if the parameter values may always be expressed using the same exponent without losing significant figures, permitting a single exponential code to be used for all values of this parameter. Thus, the following values:

		Significant digits	Exponent		Significant digits	Exponent
1	can be	0001	14	as opposed	1000	11
10	coded as:	0010	14	to:	1000	12
100		0100	14		1000	13
900		0900	14		9000	13
9009		9009	14		9009	14

Both methods of coding would be correct, and either could be used. In general practice the second method would normally be selected. Plate 6, a form for reporting and coding water quality data, illustrates how this feature of the system is handled.

If it is known that data on the same parameters will be collected for some time, and if experience has indicated the number of digits in the laboratory results are not likely to vary materially, a fixed field card may be the most efficient way of preparing these particular data for storage. In this case a storage program similar to that being used for old data from the PHS Water Pollution Surveillance System can be used. (See: *Use of Data From Other Systems* (p. 17).) This approach would be applicable particularly for intensive studies, such as the current comprehensive river basin studies of the Public Health Service. The parameter card shown in figure 6 could be used for storing data other than that which would be produced repeatedly, and which would, therefore merit being handled by the fixed field approach. For the fixed field parameter card there should be provided six columns for the station serial number and columns for the date(s) of sampling, just as in figure 6. The remaining columns (19–78) can be used in any way necessary to accommodate the fixed fields for the repetitive data, with the decimals at specific points. Thus, it will not be necessary to use the procedure for fixing decimal points prepared for the approach taken in figure 6. Data stored in this way can be retrieved by using the proper parameter code in the normal retrieval procedure.

DELETING AND CHANGING DATA

In any procedure for storing and retrieving data it is important to be able to delete or change data already stored if errors are discovered or older descriptive information must be replaced. In this system several methods of doing this have been provided:

Variable Data

1. To delete *all* data for a given station and date, columns 1–18 of the parameter card are punched as previously described and the letters D E L are coded into columns 19–21.

2. To delete an individual parameter for a given

WATER QUALITY DATA

LABORATORY BENCH DATA

(Sample)
STATION DESIGNATION

YR MO DA OF SAMPLE

HOUR & MINUTE OF SAMPLE OR LAST DATE OF COMPOSITE SAMPLE

ITEM Temperature, water UNIT °C.
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
/ 7 5

ITEM Coliforms UNIT /100 ml
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
1 5 0 0 0 0 0

ITEM BOD - 5 day UNIT mg/l.
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
2 1 6

ITEM Berylium UNIT µg/l
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
0 0 6

ITEM Phosphate - total UNIT mg/l
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
0 1

ITEM Chlorides UNIT
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6
2 7 0

ITEM UNIT
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6

ITEM UNIT
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6

ITEM UNIT
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6

ITEM UNIT
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0
8 7 6 5 4 3 2 1 0 1 2 3 4 5 6

COMPUTER CODED DATA

STATION SERIAL NUMBER (1-6) YR (7-12) MO DA (13-18)

PARAMETER CODE AND VALUE (19-23) EXPONENT (28) RMKS (29) (30)

0 0 0 1 0 | 1 7 5 0 | 1 | 2 |

(24-27)

3 1 5 2 3 (31-35)
1 5 0 0 | 1 | 7 | (36-39) (40) (41) (42)

0 0 3 1 0 (43-47)
2 1 6 0 | 1 | 2 | (48-51) (52) (53) (54)

0 1 0 1 0 (55-59)
6 0 0 0 | 0 | 1 | (60-63) (64) (65) (66)

0 0 6 5 0 (67-71)
1 0 0 0 | 0 | 0 | | | (72-75) (76) (77) (78) CHG (79)

NEXT CARD – REPEAT COLUMNS 1-18 ABOVE

0 0 9 4 0 (19-23)
2 7 0 0 | 1 | 3 | (24-27) (28) (29) (30)

(31-35)
(36-39) (40) (41) (42)

(43-47)
(48-51) (52) (53) (54)

(55-59)
(60-63) (64) (65) (66)

(67-71)
(72-75) (76) (77) (78) CHG (79)

PLATE 6

date, the parameter card is punched as if data were being stored except that a D is punched in the left hand column of the significant digit field. All fields of those cards where numbers appear in the significant digit columns are treated as valid data and the information contained in these fields will be stored. This permits the deletion of one or more individual parameter values and, if desired, the replacement of the deleted values with new figures in the same storage run. The replacement values need not be on the same cards which bear the pertinent D punches.

3. Where a relatively large amount of data must be replaced with new values, the method above is somewhat cumbersome. If column 79 of a parameter card is punched with the letter C, data for all parameters on that card, and also the data for parameters contained by all other cards for that storage run for the same station and date, will be stored in place of or merged with old data.

4. To delete the station code serial and all data for all dates associated with this station, use the following procedure:

a. Punch the station code serial in columns 1 through 6 of location card I.

b. Punch DELETE S in columns 7 through 14 of the above card.

c. Punch the number "1" in column 80 of the above card.

5. If it is desired to delete all data for all dates for a station, but retain the station code serial in the system, the following procedure is to be used:

a. Punch the station code serial in columns 1 through 6 of the parameter card.

b. Punch DELETE S in columns 7 through 14 of the parameter card.

c. Punch the number "1" in column 80 on the above card.

d. Completely recode location cards III and IV in the desired manner. It is necessary to recode cards III and IV completely even though only a portion of the information is to be changed.

2. To change all information resulting from the use of location card(s) V (the entire 792 character descriptive block) use the following procedure:

a. Punch the station code serial in columns 1 through 6 of location card I.

b. Punch the word CHANGE in columns 7 through 12 on the above card.

c. Punch the number "1" in column 80 of the above card.

d. Leave column 79 of location card V blank and punch a new descriptive paragraph as desired. The recoding information in the 792 character block can be done in conjunction with the recoding of location cards III and IV, or it can be done independently.

3. To change individual lines of information in the 792 character block while leaving the remaining information unchanged, use the following procedure:

a. Punch the station code serial in columns 1 through 6 of location card I.

b. Punch CHANGE in columns 7 through 12 of the above card.

c. Punch the number "1" in column 80 of the above card.

d. Punch the desired new information on location card V and punch column 79 of location card V with proper number to show which line is being changed. The proper numbers for punching in column 79 are: 0, to change line number 1; 1 to change line number 2, etc.; and "J" to change line number 11. When data are initially stored for the descriptive paragraph, the desired line in the printout is selected by using this same code.

Fixed Data

While data stored on location cards III, IV, and V are more or less permanent in nature the system user still must be able to change this information.

1. To change any or all information on location cards III and IV, except the agency and type of station, use the following procedure:

USE OF DATA FROM OTHER SYSTEMS

Since it is very important to be able to use data already being stored in other types of systems, a procedure has been developed for accomplishing this in the computer. Data can be accepted me-

chanically from any other system which uses a fixed field card approach for the handling of parameter values, and for which stations can be identified by a specific code. Data already produced in the PHS Water Pollution Surveillance System, for example, are being transferred this way.

PARAMETER ABBREVIATION CARD:

The parameter abbreviation card shown in figure 7, appendix III, is used to store the parameter abbreviations that are to be the column headings for the different parameters on the data printout format given by table 2 and also to locate the decimal points of the parameter values on the printout. Thus, the abbreviation card for five day BOD would be coded as:

```
COLUMN    1 2 3 4 5   6   7 8 9 10 11 12 13 14
PUNCH     00310   2     BOD

          15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
           5 DAY      MG/L
```

Each time the parameter 00310 is requested in a retrieval run,

BOD
5 DAY
MG/L

will be printed as a column heading and the BOD values will be printed under this heading.

Column 6 locates the decimal point for the values on the printout form and represents the number of characters to the right of the units digits, counting the decimal point as one unit. Thus all BOD values in the example above will be printed as: XXXX.X

Particular note should be made of the parameters 00001, 00002, and 00003 concerning sampling station location (*1*). These parameters, when used, denote the point in the stream cross section at which the sample was collected. The horizontal location may be stated in terms of feet from the right bank, or in terms of a percentage of the total width of the stream measured from the right bank, looking upstream. This means of locating sampling points within the stream cross section supplements the less flexible method shown on page 11. Parameter number 00007, "Samples in Composite," is used to show the number of individual samples which constitute the total composite, and is applicable whenever a composite is not obtained by some continuous sampling device.

As indicated earlier, special problems arise in the handling of biological data. Any one sampling may reveal the presence of very many different kinds of organisms. Furthermore, for reasons of economy, or because of varying degrees of familiarity with the different forms, biologists must frequently identify their material to different levels of the biological scale. Some forms will be reported at times to species, at other times to genera; still others may be identified only to family, or higher in the scale, depending on the worker, the time available, and the requirements of the study. Nonetheless, the biologist needs to be able to store and retrieve these data in such a way that he can get a written record of what was found in a given sample no matter how detailed the identification was. Furthermore, he may want to know how the sample was collected and he must be able to correlate the presence of any selected form with other water quality variables. These special problems, and requirements, dictate the use of a special system for the identification and handling of biological data. Such a system, to be made a part of STORET as soon as possible, is now under development.

Retrieval Routine

After the information contained on the location cards, parameter abbreviation cards and parameter cards has been stored on magnetic tape, a retrieval routine is used to obtain printouts in the formats given by tables 2 and/or 3 and/or magnetic tapes that can be used for further statistical processing.

The printout shown in table 3 contains permanent information about a station: the station serial number, the location name, the State name,

TABLE 2

STATION CODE SERIAL
LOCATION NAME
STATE
MAJOR BASIN NAME
MINOR BASIN NAME
CROSS SECTION NUMBER

DATE FROM TO	TIME OF DAY	32005 CCE UG/L	31503 COLIFORM DLY ENDO MF/100ML	00310 BOD 5 DAY MG/L	(Par. No. (Par.Abrr.)						
62 11 12	11 45		1500000	10.4							
63 01 03		98									
63 01 18	C*										
63 01 05			5K	3.2							
63 02 16	13 03		100	1.4							
63 03 08			24000	6.0							
63 04 15			3500								
					8 Digits Column						
					Remarks						

* C indicates final date of composite sample

TABLE 3

(STATION CODE SERIAL) XXXXXX
(LOCATION NAME) (32 CHARACTERS)
(STATE) (16 CHARACTERS)
(MAJOR BASIN NAME) (24 CHARACTERS)
(MINOR BASIN NAME) (40 CHARACTERS)
(CROSS SECTION NO.) X

		RIVER SYSTEM	LEVEL II	III	IV	V	VI	VII	VIII	IX	X	XI	TOTAL
STATION	INDEX	XXXXXXX	XXXXXX	XXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
	MILES	XXXX.XX	XXXX.XX	XXX.XX	XXX.XX	XXX.XX	XXX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XXXX.XX
CONTROL POINT	INDEX	XXXXXXX	XXXXXX	XXXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	XXXX	
	MILES	XXXX.XX	XXXX.XX	XXX.XX	XXX.XX	XXX.XX	XXX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XXXX.XX
STATION TO CONTROL POINT	MILES	XXXX.XX	XXXX.XX	XXX.XX	XXX.XX	XXX.XX	XXX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XX.XX	XXXX.XX

DESCRIPTION

(PARAGRAPH - 792 CHARACTERS)

the name of the major and minor basin where the station is located, or other words of identification, the station code, information pertaining to a control point, and a descriptive paragraph. The above information, except for that pertaining to a control point, is identical to the information stored by use of the location cards. The control point, to be discussed in more detail below, is a point so located on a stream that stations above and/or below it can be retrieved by the retrieval program.

Since, in any one retrieval run, there can be a number of control points, the location code of the control point above or below which the station is located and also the mileage from the control point to the station is given in this printout.

Example:

If we consider station 5 in the hypothetical river system shown in plates 2 and 3, to be a control point and station 3 is retrieved under control of this point, then the printout shown in table 4 will be given for station 3. The descriptive paragraph is an example of what might be included in the description of station 3, though considerably more space is available than used in this example.

The water quality data for an individual sampling station is printed in the format given by table 2. This type printout locates each value vertically under its abbreviation and horizontally from the date and time of the sample. The decimal points for all values of a single parameter are located in the same column. With the exception of those which are whole numbers, values are limited to seven decimal digits; eight decimal digits are available if values are whole numbers, with no decimal point showing.

The remark code which is stored with the parameter value will be converted to a letter code by the retrieval program and printed to the right of the individual values. Letters are used instead of numbers, since this code must be distinguishable from the parameter value. If no remark has been coded, the remark column will remain blank.

Up to 50 parameters can be obtained per station per run of the retrieval routine. Since up to 10 parameters per sheet can be printed in the format given in table 2, up to 5 different combinations of parameters can be printed per station.

PARAMETER CONTROL CARD:

Control of parameters to be retrieved will be accomplished by use of the parameter control card (figure 8, appendix III). In this card the code numbers of the parameters to be retrieved are coded in fields I–X in the same order that they are to be printed on the printout given in table 2. If certain fields are to be left blank in the printout, 99999 is coded in this field. Thus, if the parameters 10500–10506 and 11000 are to be retrieved, this card would be coded, starting with column 2 as:

COLUMN PUNCH	2 3 4 5 6	7 8 9 10 11	12 13 14 15 16
	10500	10501	10502
17 18 19 20 21	22 23 24 25 26	27 28 29 30 31	32 33 34 35 36
10503	10504	10505	10506
37 38 39 40 41	42 43 44 45 46	47 48 49 50 51	
99999	11000	99999	

Since up to 50 parameters can be retrieved in one retrieval run, up to 5 cards of this type will be needed per run.

STATION CONTROL CARDS:

The retrieval program is designed to retrieve information for specific individual stations or for groups of stations.

If individual stations are to be retrieved, control card I, figure 9, appendix III, will be used. The 6 digit serial codes of the stations to be retrieved are punched in columns 1–72 of this card. Up to 48 individual stations can be retrieved in one run, requiring a maximum of four control cards.

If groups of stations are to be retrieved, then control cards II, III, and IV are used (see appendix III, figures 10, 11, and 12).

TABLE 4

(STATION CODE SERIAL) 380003
(LOCATION NAME) LOCATION NAME OF STATION 3
(STATE) STATE IN WHICH STATION 3 IS LOCATED
(MAJOR BASIN NAME) MAJOR BASIN IN WHICH STATION 3 IS LOCATED
(MINOR BASIN NAME) MINOR BASIN IN WHICH STATION 3 IS LOCATED
(CROSS SECTION NO.) 4

			RIVER SYSTEM	II	III	IV	V	VI	VII	VIII	IX	X	XI	TOTAL
	STATION	INDEX	5110003	000020	00000	0000	0000	0000	0000	0000	0000	0000	0000	
		MILES	0181.25	0121.00	000.00	000.00	000.00	000.00	00.00	00.00	00.00	00.00	00.00	302.25
23	CONTROL POINT	INDEX	5110003	000020	00010	0020	0000	0000	0000	0000	0000	0000	0000	
		MILES	0181.25	0178.00	100.02	041.11	000.00	000.00	00.00	00.00	00.00	00.00	00.00	500.36
	STATION TO CONTROL POINT	MILES	0000.00	0057.00	100.02	041.11	000.00	000.00	00.00	00.00	00.00	00.00	00.00	198.13

DESCRIPTION

Long. 86D37M42S, Lat. 39D17M41S, 100 ft. below Mills Bridge on Route 11

Flat swampy land - no industry

Control card II is used when all stations *above* a certain point, called a control point, are to be retrieved. The control point is located on a suitable map and then coded just as a station at the same point would be coded, except that the mileage on the last level only need be obtained.* The index codes of the point are then punched into columns 7–56 of the control card II and the mileage on the last level in columns 57–62. The number of the level on which the point is located is punched into columns 63 and 64. Columns 1–6 are left blank unless the programer wishes to specify that only stations in a given State, or under a specific agency, are to be retrieved. If the programer wishes to select on a State code, the State code is punched in columns 1 and 2. In order to select on an agency code the six-digit agency code is punched in columns 1–6.

Example:

To select all stations above station 5 in hypothetic basin 51, control card II will be coded as follows:

COLUMN	7 8 9 10	11 12 13	14 15 16 17 18 19
PUNCH	5110	003	000020

20 21 22 23 24	25 26 27 28	57 58 59 60 61 62	63 64
00010	0020	004111	04

Control cards III and IV are designed to be used when stations located below a control point, or when all stations located on a tributary that intersects a stream below a control point, are to be retrieved. They can, however, also be used to control the retrieval routine so that stations above a control point will be selected.

Columns 1–6 of control card III are normally left blank. However, if the programer wishes to select on an agency code, the 6 digit agency code is punched in columns 1–6. Columns 7–78 of control card III and columns 14–45 of control card IV are punched as if the control point were actually a station. Column 71 in control card IV is coded with a number to tell the computer what stations to select with respect to the control point. Specifically, column 71 is coded with a:

[1] If all stations above the control point are to be selected.

[2] If stations on segments of rivers which are downstream from the control point are to be selected.

[3] If stations on segments of rivers which are downstream from, and all stations above the control point are to be selected.

[4] If stations below and on tributaries that intersect the stream below the control point are to be selected.

[5] If stations above the control point and also the stations below and on tributaries that intersect the stream below the control point are to be selected.

Columns 72–79 in control card IV are left blank if column 71 is coded with a 1. In all other cases, however, these columns are used to tell how far below the control point the stations are to be selected; i.e., a point somewhere below the control point is selected such that no stations are to be retrieved that are below this point. Since this second point is below the first control point it will have the same level coding as the control point up to and including the index code of the level on which the second point is located. Therefore, this second point is completely located by coding into columns 72 and 73 the number of the level on which the second point is located and in columns 74–79, the miles to this second point on the last level. If *all* stations below the control point are to be retrieved, the second point becomes the mouth of the base river. Thus, column 71 of control card IV should be coded 2 and columns 72 and 73 should be coded 01, since the base river is a first level stream. Zeros will be coded into columns 74–79, to represent zero mileage.

*The other mileages are not needed since the retrieval program selects stations by comparing the index codes up to the level of the control point and only then does it compare the mileage of a level.

Example:

If we wish to retrieve all stations above station 8 and also between station 3 and station 8 in river systems shown in plate 3, the following code will be punched into control cards III and IV.

CONTROL CARD III

	7 8 9 10	11 12 13	14 15 16 17 18 19		
COLUMN PUNCH	5110	003	018125		
20 21 22 23 24 25	26 27 28 29 30 31	32 33 34 35 36	37 38 39 40 41		
000020	017800	00010	10002		
42 43 44 45	46 47 48 49 50	51 52 53 54	55 56 57 58 59	77 78	80
0020	14895	0030	12050	05	3

CONTROL CARD IV

	71	72 73	74 75 76 77 78 79	80
COLUMN PUNCH	3	02	012100	4

Using these control cards, stations 3, 5, 8, 9, 10, 11, 12, 13, and 14 would be retrieved.

If, however, column 71 in control card IV were coded with:

[1] Stations 8, 9, 10, 11, 12, 13, and 14 would be retrieved.

[2] Stations 3, 5, and 8 would be retrieved.

[4] Stations 3, 5, 7, and 8 would be retrieved.

[5] Stations 3, 5, 7, 8, 9, 10, 11, 12, 13, and 14 would be retrieved.

Although the different control cards have been discussed individually, all three types can be used in one retrieval run.

The actual river system shown in plate 7 will illustrate further the method of using control points for the retrieval of selected groups of stations. In this river system, if point "4" is used as a control point and point "3" as the second point, and then the stations located below the control point and above the second point are retrieved (2 punched in column 71 of control card IV) all stations which are located on the waterways indicated by the dashed line connecting points "3" and "4" will be retrieved. This type of retrieval could be used to obtain water quality data for stations located below a sewage outfall. If, however, point "3" is used as a control point and point "1" as the second point, and all stations located below and on tributaries below are retrieved (4 punched in column 71 of control card IV) all stations located in the area enclosed by the dot-small dash line will be retrieved. This type of retrieval will be used when the programer wishes to retrieve all stations located in a minor basin.

A third possibility offered by the system, i.e., the retrieval of all stations above a control point, is illustrated if point "2" is used as a control point and all stations located above this point are retrieved. (One punched in column 71 of control card IV.) In this case, all stations in the small area bounded on the west by the dot-long dash line will be retrieved. This type of retrieval could be used to retrieve all sewage outfalls located above a particular point.

DATE CONTROL CARD:

The final card needed to control the retrieval routine is the date control card (figure 13, appendix III.) This card is used to control the interval of time for which the parameters are to be retrieved, and also the type of station to be retrieved. Thus, for example, if the parameters are to be retrieved for the year 1959, columns 2–13 of this card will be punched 590000 and 591231. Columns 18–20 will be coded with the following characters for the various station types:

002—Stations, other than water intakes, where water quality data are collected.

004—Stations where stream flow is recorded.

010—Stations which are municipal intakes where data on raw water are collected.

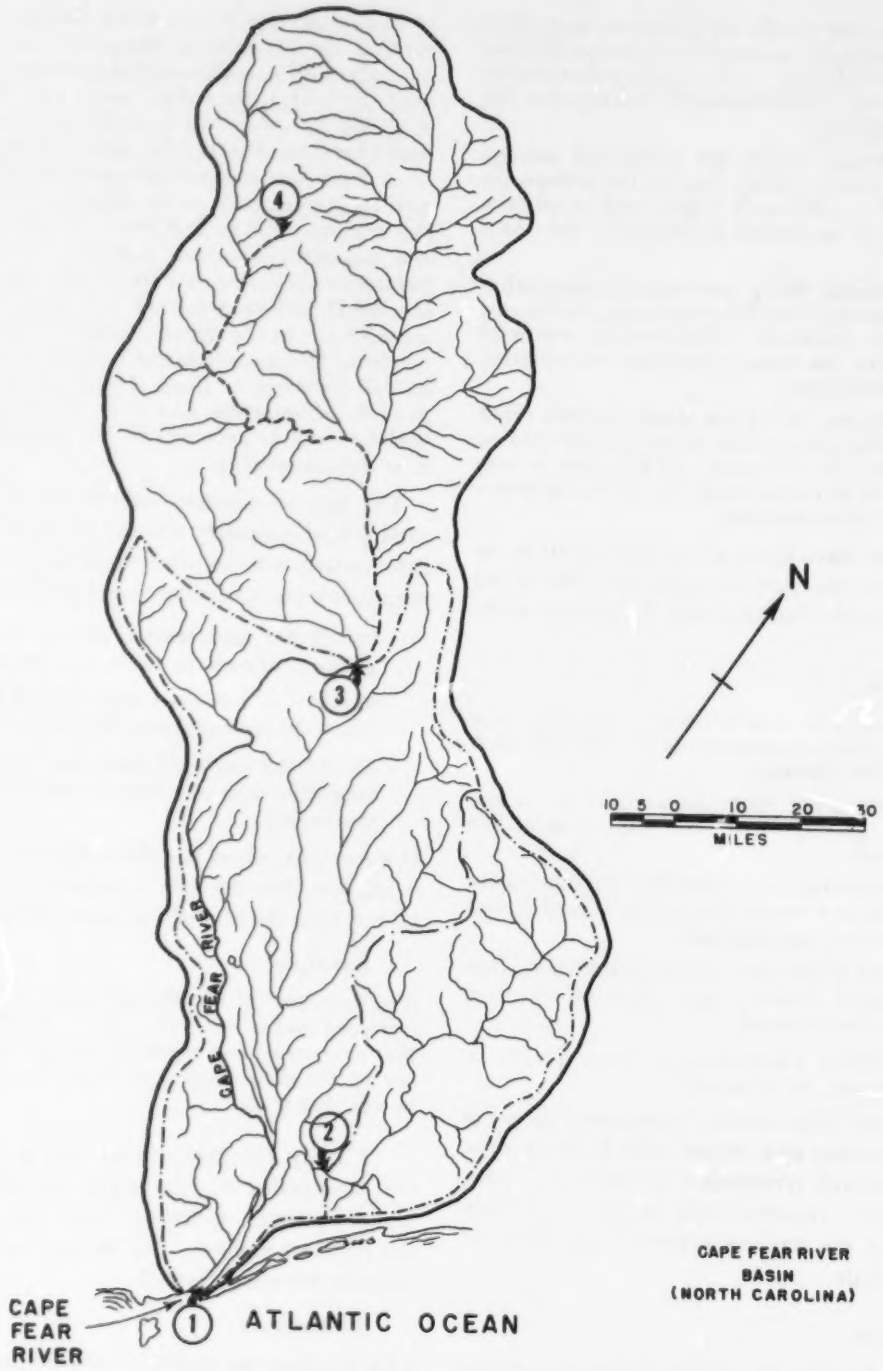

4
3
2
1
CAPE FEAR RIVER
N
10 5 0 10 20 30
MILES
CAPE FEAR RIVER
CAPE
FEAR
RIVER
ATLANTIC OCEAN
CAPE FEAR RIVER
BASIN
(NORTH CAROLINA)

PLATE 7

020—Stations which are intakes to municipal water treatment plants where data on finished water are collected. (This code is also used to retrieve the water treatment information for these stations.)

040—Stations which are municipal sewage outfalls where quality data on the sewage are collected. (This code is also used to retrieve the sewage treatment information for these stations.)

100—Stations which are water intakes other than municipal intakes where data on the raw water are collected. (This code is also used to retrieve the water treatment information for these stations.)

200—Stations which are waste outfalls other than municipal outfalls where quality data on the waste are collected. (This code is also used to retrieve the waste treatment information for these stations.)

If two or more types of stations are to be retrieved in one run, then the type code will be the sum of the codes characterizing the stations to be retrieved.

Example:

If all stations where data on water quality in a stream have been recorded are to be retrieved, then the code will be 112 since:

002—Stations other than water intakes where water quality data on the stream are collected.

010—Stations which are municipal intakes where data on raw water (i.e., water directly from the stream) are collected.

100—Stations which are water intakes other than municipal intakes where data on the raw water are collected.

112—All stations where data on water quality in the stream are collected.

If an individual station is requested by using the serial number and control card I, this station will be retrieved irrespective of its type. Also, any parameter requested will be extracted from the data for the stations requested regardless of the station type.

Example:

If Station A is a water quality station on a stream and it is necessary to retrieve stream quality data at station A, and also data on discharges for all outfalls above station A, the parameters to be retrieved would be coded into the parameter control card (or cards), the 6 digit serial code for station A would be coded into control card I, the index code of station A would be coded into control card II, and the date control card would be coded. The type of station code in the date control card for this example will be 240, to cover both municipal and nonmunicipal waste outfalls. If one of the parameters required was BOD, for example, then the BOD values retrieved for station A itself would refer to the stream, while the BOD values retrieved for the outfall stations would refer to the outfall discharges. Each station would be identified by information like that shown on table 4 which would be printed out just before the pertinent data on quality.

The *type of printout* field in the date control card is used to specify whether or not the retrieved information is to be printed out and/or stored on magnetic tape. Columns 14 and 15 are coded with:

01—If for each station that is retrieved the printout shown in table 3 is to be printed.

02—If the retrieved data are to be printed out in the format shown in table 2.

04—If the retrieved data are to be stored on magnetic tape in a format useful for further processing.

If a combination of the above formats is required in one run, then the code to be punched in columns 14 and 15 is the sum of the codes of these formats.

Example:

If a printout of the descriptions of the stations retrieved and also a data tape to be used for further processing are required, but no actual printout of the data itself is needed, then the code to be punched is 05.

Finally, the date control card is used to control the number of originals of the printout to be made per run. Columns 16–17 are coded with the required number. Up to 8 carbon copies can be made for each original.

Example:

If 2 originals are needed, then columns 16–17 will be coded with an 02.

References

(1) *Parameter Code List for the STORET System*, Basic Data Branch, Division of Water Supply and Pollution Control, Public Health Service, U.S. Dept. of Health, Education and Welfare, Washington, D.C., 20201, March 1964.

(2) *Location Coding for the STORET System*, Basic Data Branch, Division of Water Supply and Pollution Control, Public Health Service, U.S. Dept. of Health, Education, and Welfare, Washington, D.C., 20201, March 1964.

Appendix I

STORAGE OF DATA

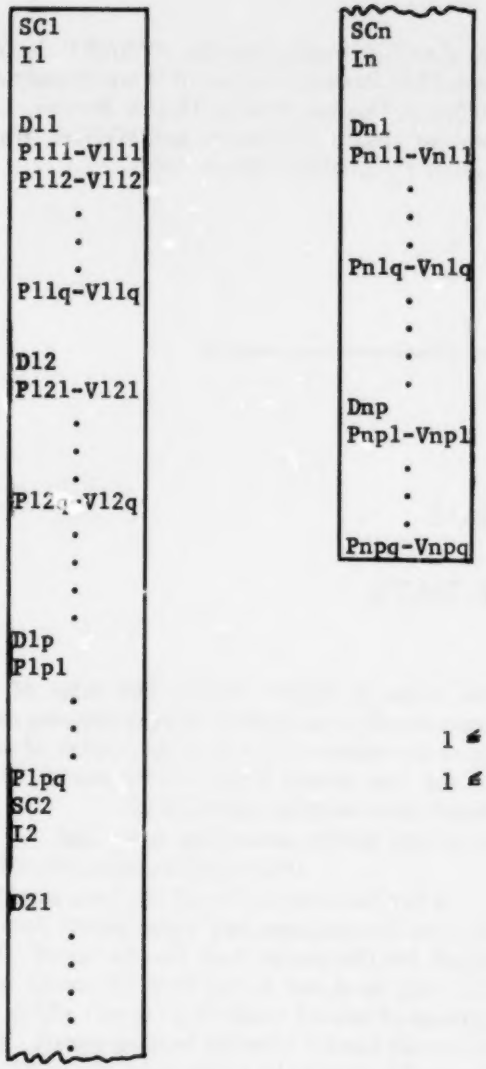
All of the information to be contained in this system is stored on magnetic tape. Although a number of different tapes are used by the storage and retrieval programs, the most important tape is the data-location tape, for this tape contains the information that will be retrieved by the system. A description of this tape and its uses will show how the data are stored and retrieved.

The information contained by the data-location tape is stored in hydrologic sequence based on the location code. (See location cards I and II.) After this code has been found for a station it is stored on the data-location tape together with descriptive information about the station in an individual record called the location record, a record being a related group of words, separated by gaps in the tape from other records. (See discussion of location cards III, IV, and V.) This record is located on the tape, after storage, in a position directly after the records for the stations having location codes of lower value and directly before the records for the stations having location codes of higher value. The value of the location code is the magnitude of the number made up of the indexes followed by the number of miles on the last stream level. Thus, station 12 in plate 3, has a location code value of:

51 10 003 000020 00010 0020 0030 0030
0020 0010 0020 0000 0000 007999

After the location record has been stored on the data-location tape, any water quality data obtained for this station may also be stored. This data will be stored in the form of record items (groups of related words in a record) which will be located directly after the location record. Each item will be headed by a date code (the coded date of the sample) and will contain the values of the parameters measured on that date, each parameter value being accompanied by the parameter number. (See discussion on parameter card.) The items will be stored by date order and the parameter values by parameter number order within the items. Thus, the final data-location tape will be in the logical order shown in diagram I.

DIAGRAM I



Where:

$VSC1 < VSC2 < \ldots < VSCn$

$1 \leq i \leq n$ $\quad Di1 < Dia < \ldots < Dip$

$1 \leq j \leq p$ $\quad Pij1 < Pij2 < \ldots < Pijq$

DEFINITIONS:

SCi = i'th station location code

Ii = descriptive information on station i

VSCi = value of i'th station location code

Dij = date of j'th sample at station i

Pijk = parameter number of value Vijk for date Dij at station i

Appendix II

RETRIEVAL OF DATA

The retrieval routine will retrieve information from the data-location tape by the following procedure.

By using control cards (see discussion of "Retrieval Routine"), a programer will tell the computer what information is to be retrieved, for which stations, and how the information is to be processed and printed.

The computer, under the control of these cards, then reads the data-location tape and when it comes to a location record tests to see if any information from this station is to be retrieved. If the test is negative, i.e., no information is to be retrieved from this station, the computer continues to read the tape until the next station location record is reached and again it performs the test. This procedure is repeated until a station is reached where the test is positive. The information from the location record of this station is then processed and transferred within the computer to a second tape called the print tape.

The computer then continues reading the data-location tape until it comes to a record item containing parameter values. The date code of this item is then tested to see if it is within a specified interval of time. (See discussion of control cards.) If the date is not within the interval, the computer continues reading and testing the dates of the items.

If the date is within an interval the parameter numbers in the record item are then tested to see if any coincide with the parameter numbers requested. (See discussion of control cards.) If none coincide, the computer checks the next item to see if the date is within the specified time interval. If, however, one or more parameter numbers do coincide, the parameter values stored with these parameter numbers are processed and stored on the print tape. If it were requested that these values also be stored on a third tape to be used for further processing, the computer also performs this operation at this time. After this processing is completed, the computer goes to the next date and test. When a date is found to be greater than the last date in the prescribed interval of time, or when a location record is found, the computer reverts back to testing the station codes.

When the point on the tape is reached at which no more stations remain where the location test is positive, the computer rewinds all the tapes and then by use of print tape prints out the information that was retrieved.

Appendix III

CARD FORMAT AND CODING

FIGURE 1

Location Card I

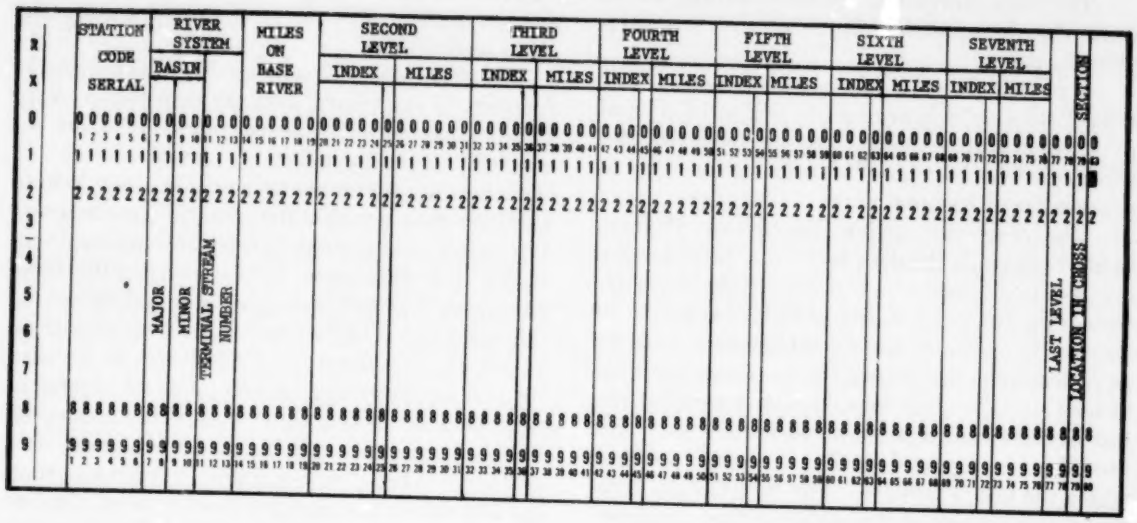


FIGURE 2

Location Card II

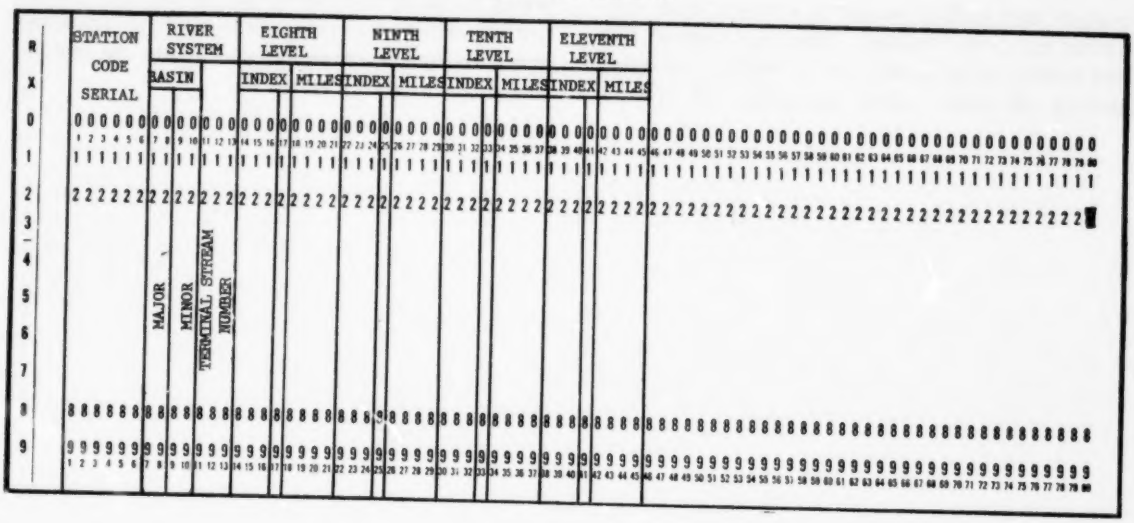


FIGURE 3

Location Card III

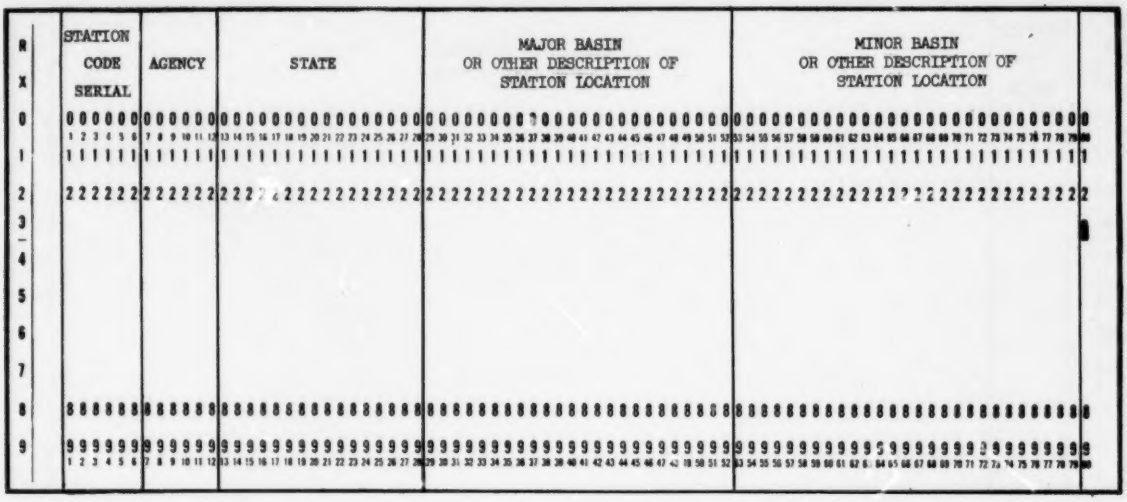

STATION CODE SERIAL
AGENCY
STATE
MAJOR BASIN OR OTHER DESCRIPTION OF STATION LOCATION
MINOR BASIN OR OTHER DESCRIPTION OF STATION LOCATION

FIGURE 4

Location Card IV

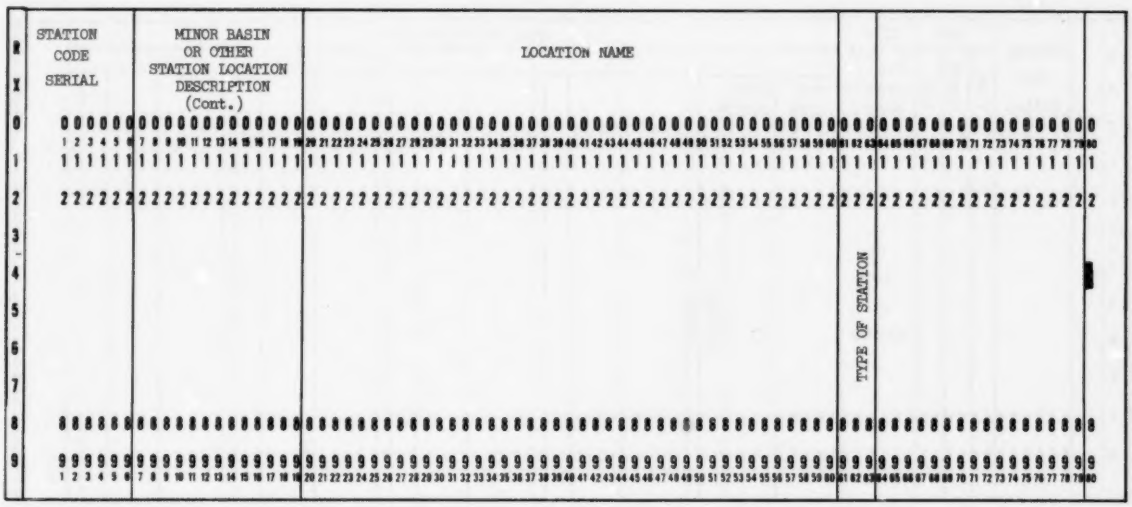


FIGURE 5

Location Card V

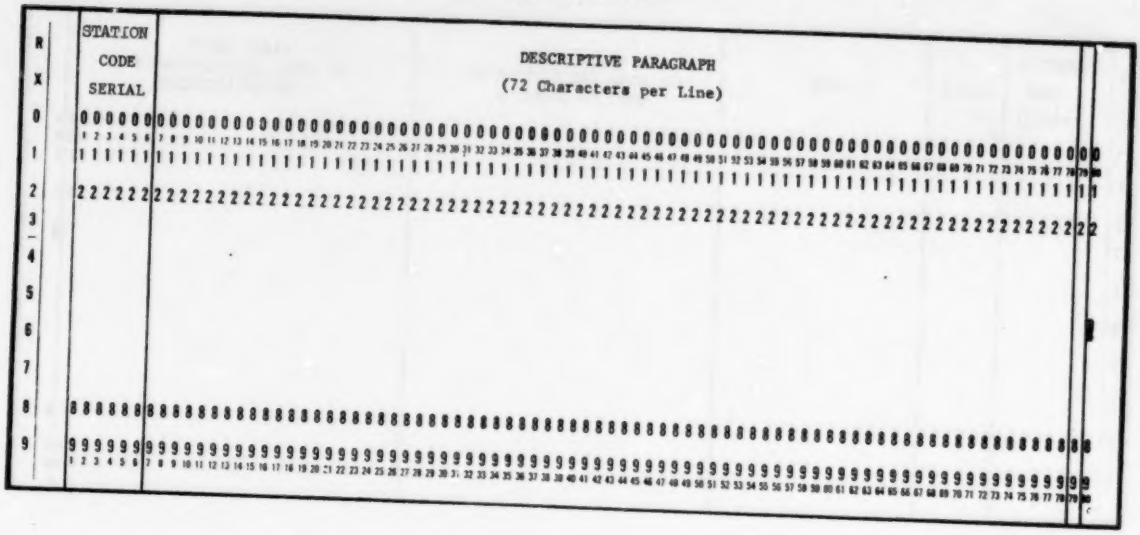


FIGURE 6

Parameter Card

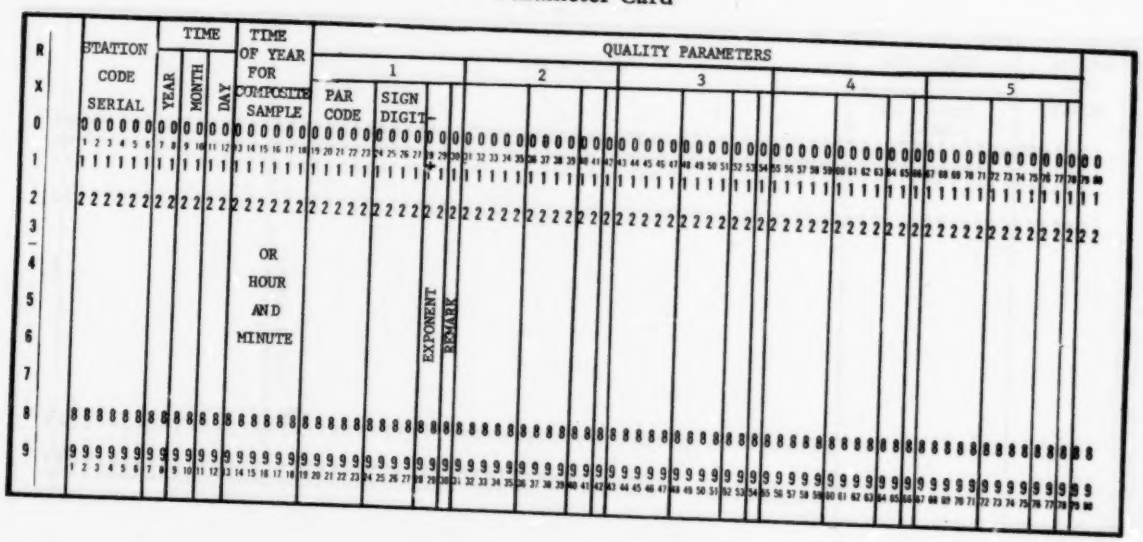


FIGURE 7

Parameter Abbreviation Card

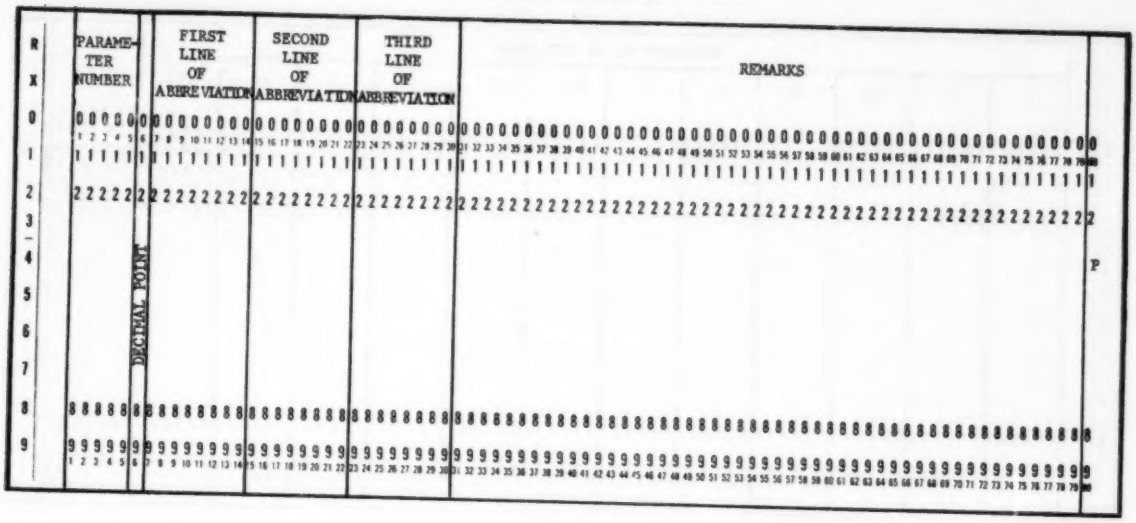


FIGURE 8

Parameter Control Card

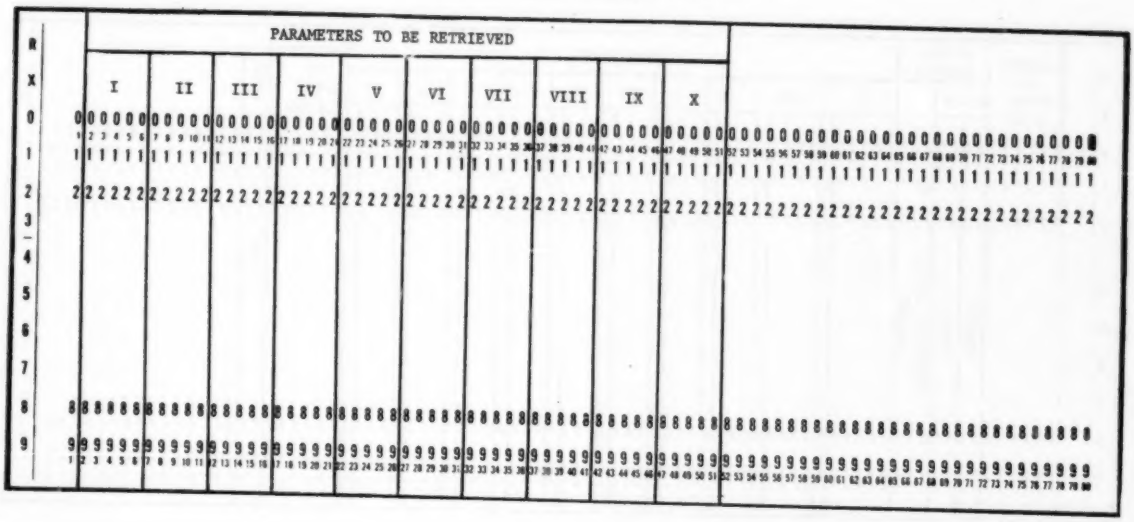


FIGURE 9

Control Card I

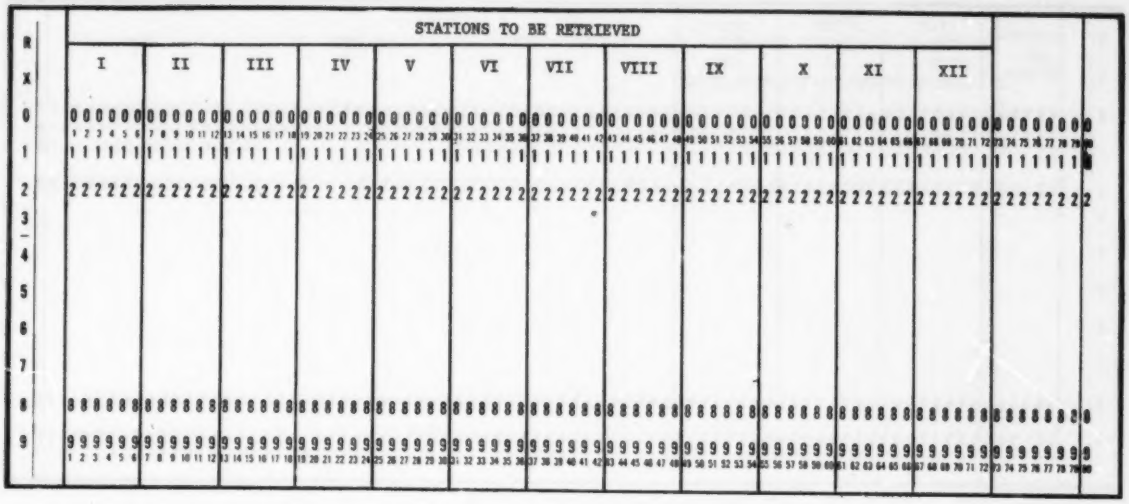


FIGURE 10

Control Card II

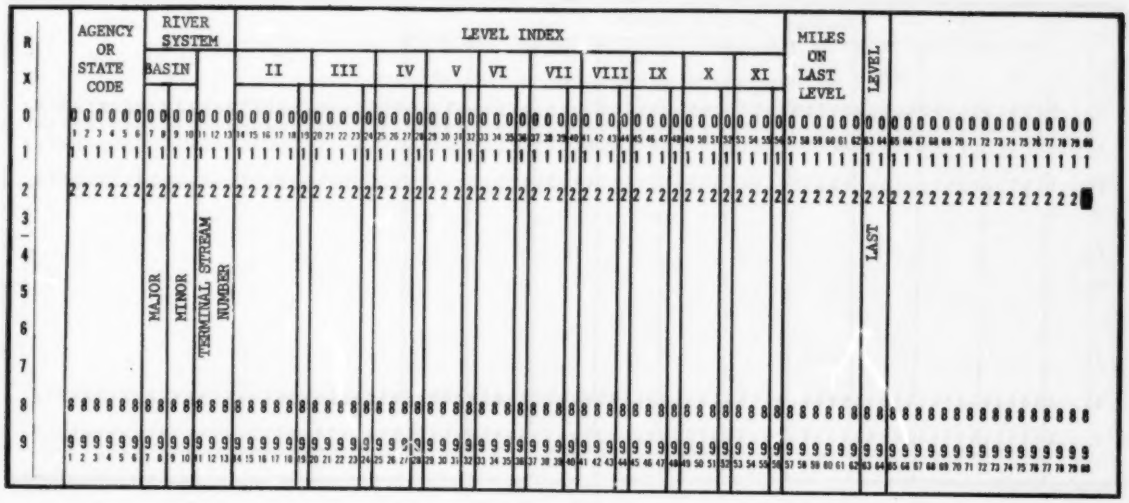


FIGURE 11

Control Card III

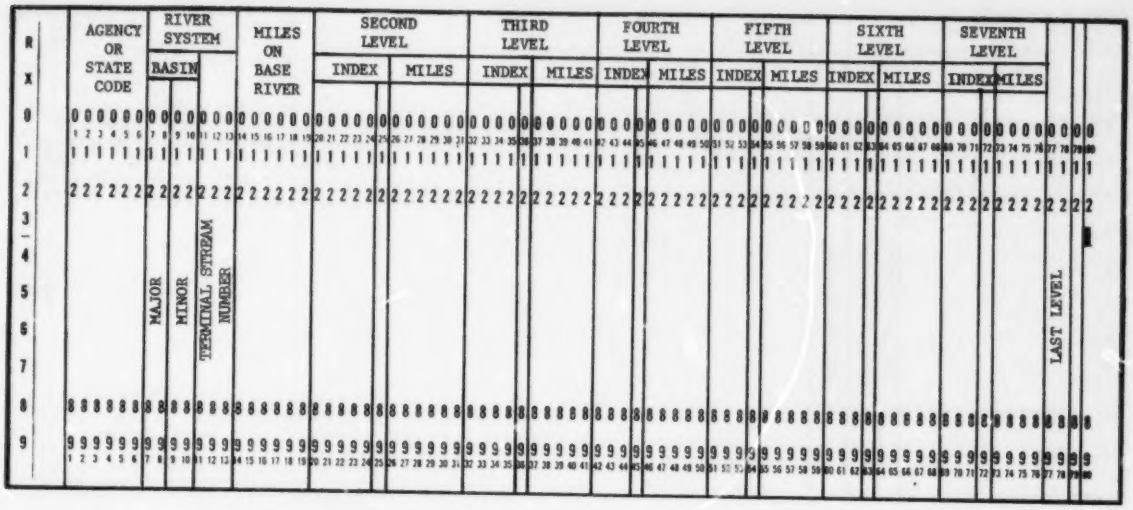


FIGURE 12

Control Card IV

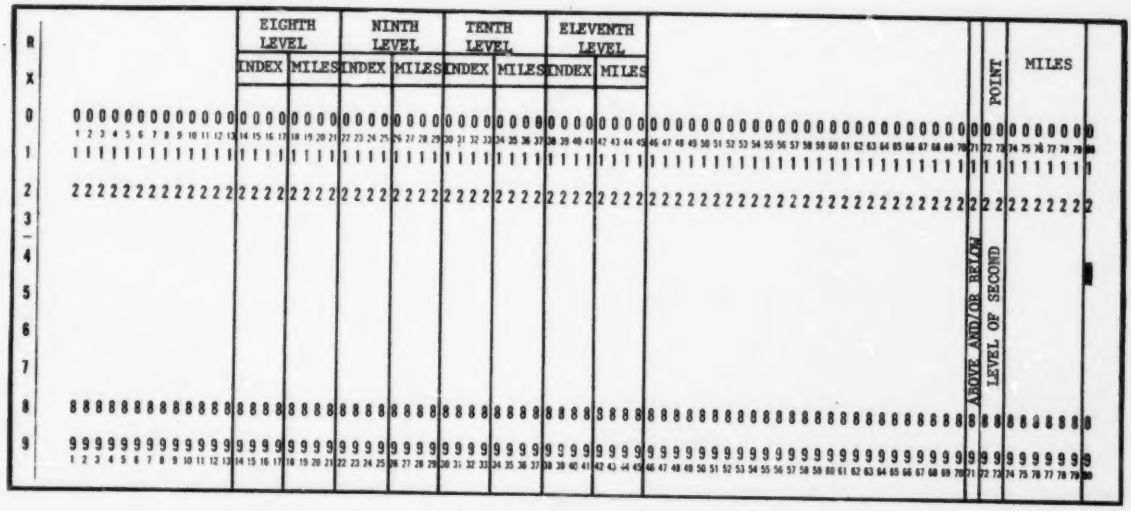
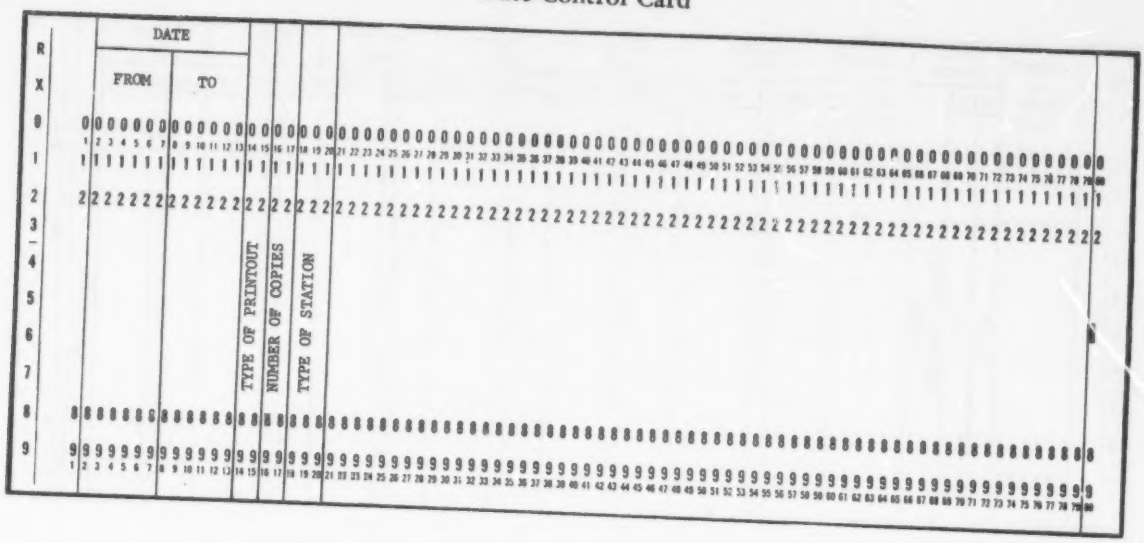


FIGURE 13

Date Control Card



U. S. GOVERNMENT PRINTING OFFICE : 1964 O - 730-416